Twitter Thread by Pratham Prasoon



Pratham Prasoon
@PrasoonPratham



Which libraries do you really need to get started with Machine Learning and why?

■■

First and foremost, make sure that you have nailed the fundamental concepts of Python because machine learning requires a lot of programming!

(2 / 19)

If you know these topics, then you are good to go with machine learning in Python

- Object-oriented programming in Python:Classes,Objects,Methods
- Lists & List functions
- List comprehension
- List slicing
- String formatting
- List,Dictionaries & Tuples

(3 / 19)

Now, let's understand what popular python machine learning libraries do.

We will talk about■
- TensorFlow (+ Keras)
- PyTorch
- Pandas
- Numpy
- Matplotlib
- SciKit Learn
- Seaborn

(4 / 19)

Now let's cover the libraries that you have to learn (at least the basics) for machine learning

1. Pandas

Pandas is a python library that allows you to store and read data from spreadsheets (.csv, .xlsv files) in structures called Dataframes.

(5 / 19)

Dataframes allow you to work with data present in a spreadsheet.

Pandas help you make the data frame itself.

(6 / 19)

2. Numpy

Numpy allows you to manipulate the data. It replaces python lists and does the same things, like list slicing for example. However numpy lists are much faster to execute than the default python lists.

(7 / 19)

3. Matplotlib

Matplotlib is a library for plotting data into pie charts, bar charts, and whatever kinds of graphs you can imagine.

(8 / 19)

There is another library based on Matplotlib which is called Seaborn.

Seaborn is based on Matplotlib and allows you to visualize data with support for themes (as in color schemes like VS code themes) and more visualization options.

(9 / 19)

It is definitely not mandatory to learn Seaborn at the beginning of your machine learning journey.

Use it when you need to.

(10 / 19)

Why is it important to learn these libraries?

In machine learning, you will have to work with a lot of messy data! A lot!
These libraries are essential for you so that you can manipulate and analyze data.

(11 / 19)

You must always remember that data largely dictates how well your machine learning model will perform.

Do not ignore data analysis and cleaning.

It is even more important than neural network!

(12 / 19)

Now we are left with TensorFlow, PyTorch, and Scikit Learn.

Let's talk about them.

(13 / 19)

TensorFlow allows you to make neural networks, anything from computer vision to natural language processing, TensorFlow will help you out. Keras allows you to make machine learning models and is heavily integrated with Tensorflow v2 (the latest version of TensorFlow).

(14 / 19)

- PyTorch does the same things as TensorFlow in a different syntax.

- Both PyTorch and TensorFlow are equally amazing libraries.

(15 / 19)

SciKit Learn

Scikit learn does a lot of things, from regression to classification, you name it.
It is a great tool to have when working on machine learning.

(16 / 19)

Let's summarize.

Step 1: Learn Python well.
Step 2: Learn the basics of Numpy, Pandas, and matplotlib.
Step 3: Learn either PyTorch or TensorFlow or SciKit learn at the start.

(17 / 19)

Step 4: Do a project from Kaggle where you can use these tools. (Pro-tip, look at notebooks and solutions from other people, this will help you in understanding how experienced people tackle certain problems)

(18 / 19)

The best tutorials to learn these would be from the freecodecamp YouTube channel, simply search for `framework name` freecodecamp.

They have some of the best machine learning tutorials out there.

(19 / 19)